DOCUMENT RESUME

ED 203 739 HE 013 989

AUTHOR Baird, Leonard L.
TITLE The Relationship Between Ratings of Graduate Departments and Faculty Publication Rates.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-80-27
PUB DATE Oct 80
NOTE 26p.
AVAILABLE FROM Educational Testing Service, Princeton, NJ 08541.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Alumni; Chemistry; *College Faculty; *Departments; Educational Quality; Evaluation Criteria; Graduate Students; Higher Education; History; Measurement Techniques; *Productivity; Psychology; *Publications; Recognition (Achievement); *Reputation; *Scholarship; Student Attitudes; Teacher Attitudes; Teacher Effectiveness
IDENTIFIERS *Faculty Publishing

ABSTRACT

The relationship between reputational ratings of graduate-level chemistry, history, and psychology departments and their faculties' scholarly productivity was examined using data from a national sample of departments in each field. Faculty reported the number of their professional articles in journals, chapters in books, scholarly book reviews, books, and monographs. This information was requested separately for the entire career and for the last three years. Among other ratings, faculty were asked to rate the quality of the faculty in the other departments in their field. Additionally, graduate students and alumni from the program were surveyed, using various measures. Departments were divided into three groups based on their reputational ratings, and the productivity of faculty in these departments was assessed. For measures covering the entire career, the high-rated departments produced approximately twice as much per faculty member as the low rated departments; however, this trend was not evidenced for the measures over the last three years. Correlations of departmental mean publication indices with the mean reputational ratings generally suggest a moderate relationship between reputation and productivity. Scatterplots demonstrated that the ratings were not good indicators of the productivity of individual departments. The results also indicate that any particular publication measure should not be used as a measure of overall departmental quality (some departments that were high on book production were low on journal production). Additionally, both reputational ratings and publication figures were found to be unrelated to the quality of teaching as reported by students, alumni ratings, and chemistry and history faculties' concern for students. (SW)

RESEARCH REPORT

THE RELATIONSHIP BETWEEN RATINGS OF GRADUATE DEPARTMENTS AND FACULTY PUBLICATION RATES

Leonard L. Baird

Educational Testing Service
Princeton, New Jersey
October 1980

The Relationship Between Ratings of Graduate Departments and Faculty Publication Rates

Leonard L. Baird

Educational Testing Service
Princeton, New Jersey 08541
October 1980

Abstract

The relationship between reputational ratings of graduate chemistry, history, and psychology departments and their faculties' scholarly productivity was examined using data from a national sample of departments in each field. Six measures of productivity, three for the entire career, and three for the most recent three years, were related to ratings of the "quality" of the faculty, which were obtained by following procedures used by Roose and Andersen (1970). Although some statistically significant relationships were obtained, a close examination of the data indicated that ratings are inadequate assessments of the scholarly contribution of faculty or the "quality" of departments.

The Relationship Between Ratings of Graduate Departments and Faculty Publication Rates

There has been a long controversy about the relationship between a department's scholarly productivity and its reputation within its field. Some researchers have reported a substantial positive relationship--i.e., the more the faculty in a department publish, the higher its standing with other faculty in the field--thus promoting the idea that these are mutually supportive indicators of the scholarly contribution and "quality" of the department. Certainly scholarly contributions have been of primary importance in the evaluation of research, graduate training, and quality of faculty in schools and departments in universities. Publications are evaluated to provide evidence for quality of scholarship both for individuals and departments (see Jones, 1980, for an excellent review of the issues involved in assessing scholarship through publications). A number of studies have found substantial correlations between publication activity and other indicators of scholarship (e.g., Clark, Hartnett, & Baird, 1976; Hagstrom, 1971). However, some have argued that publication rates are inadequate as measures of scholarship, and that undue emphasis on sheer high publication rates is ultimately damaging to quality. More importantly, critics argue that high publication rates are unrelated or even negatively related to quality of teaching and overall contribution to the discipline through professional activities. That is, a faculty member may be an excellent professor and member of the profession without having a high output of articles, and a department may be an

This study was supported by the Graduate Record Examinations Board.

excellent place to work and study without having a high average publication rate. Conversely, a faculty member who feels under pressure to "publish or perish" may neglect his or her teaching duties and professional commitments to concentrate on publishing.

Although publication rates have been widely criticized as indicators of departmental quality, there has been even greater controversy over the meaning and utility of reputational ratings of the "quality" of graduate departments (see Clark, 1976, for a summary of the issues). Doubts have been expressed about their significance for different kinds of programs, their accuracy, their relation to graduate students' learning experiences, etc. However, the high degree of relationship between scholarly productivity and reputational ratings have been used to defend the use of each as indicators of program quality; i.e., the high relationship with publication rates has been used to suggest that reputational ratings were valid, and the high relationship with reputational ratings to show that publication rates are a legitimate measure of quality. Thus the degree of relationship between these two variables becomes critically important. Are they really highly related, and can they be used as indicators of overall "quality"?

Earlier studies had found substantial relationships between the variables but had not controlled for size; thus the larger and usually more prestigious departments tended to appear more productive. In contrast, in a recent article and a series of comments in the American Psychologist centering on the relationship of reputational

ratings with the scholarly productivity of faculty, Cox and Catt (1977) presented data that seemed to show very little relationship between the Roose and Andersen (1970) ratings of psychology departments and the productivity of their faculty as measured by the average number of articles after adjusting for the size of the department's faculty in 13 APA journals over a six-year period. Robey (1979) obtained very similar results in a study of six political science journals. The Cox and Catt study would seem to call into question the assumed relationship between scholarly productivity and ratings. However, their study was criticized on the grounds that they had: (1) misidentified faculty members in specific departments thus making their productivity measures erroneous (Kleinmuntz, 1978); (2) included many non-psychology faculty in the counts for some departments, thus making large multiuniversities appear more productive (Levin et al., 1978); (3) used the wrong number of faculty for departments, which makes departments that have recently expanded look less productive (Ross, 1978); and (4) neglected the fact that many psychology faculty publish in other APA journals and non-APA journals, and many concentrate on books (D'Amato, 1978). A fifth potential criticism is that some prestigious departments lay their claim to eminence on the substantial contributions of their faculty over their entire careers rather than their recent journal output, and a sixth is that the ratings used were not contemporary with the publications data, thus introducing other possible errors. Finally, and perhaps most important, the Cox and Catt study, as well as

virtually all others, did not include any measures of the educational quality of the departments as places to study or to work.

The purpose of this paper is to examine the question of the relationship between reputational ratings and productivity and the relationship of these to other indicators of the educational quality of departments by analyzing a set of data from a national project designed to assess the dimensions of quality in graduate education (Clark, Hartnett, & Baird, 1976). These data meet the seven criticisms just outlined by: (1) obtaining data about productivity directly from faculty members, thus eliminating misidentification; (2) including only bona fide faculty members of departments; (3) using only the N's of respondents, thus making the base N's for averages more accurate (the response rate was high); (4) including all journal publications, not just those in a specific set of journals, and also obtaining data on book and monograph publications; (5) obtaining publication figures for each faculty member's entire career, as well as the most recent period; (6) obtaining ratings contemporaneously with the publication data; and (7) including a variety of measures of the quality of the department's environment for learning and working. In addition, the data comes from three fields, chemistry, history, and psychology so that the patterns can be compared across fields.

Method

A. Faculty Samples and Variables. The basic sample consists of 511 chemistry, 584 history, and 598 psychology faculty members in

a national sample of doctoral programs. There were 24 chemistry departments, 25 history departments, and 24 psychology departments. The departments were selected to constitute a heterogenous sample in terms of size, location, earlier reputational ratings, and emphasis. (See Clark, Hartnett, & Baird, 1976, for details.)

Questionnaires concerning faculty activities and views were developed by the authors and distributed by departmental chairpersons. (The average response rate was 75 percent). Part of the questionnaire asked faculty to report the number of their: (1) professional articles in journals and chapters in books, (2) scholarly book reviews, and (3) books and monographs. This information was requested separately for the entire career and the last three years. The validity of this self-report information is suggested by an analysis of the data by Clark and Centra (forthcoming) showing that the self-reports of the last three year journal production by young psychologists correlated .85 with counts based on listings in Psychological Abstracts, which, of course, does not cover non-psychology journals.

The following six measures were used to assess productivity:

1. Articles and book chapters. total for the entire career.
2. Articles, book chapters and book reviews, total for the entire career.
3. Books and monographs, total for entire career.
4. Articles and book chapters, last three years.
5. Articles, book chapters and book reviews, last three years.
6. Books and monographs, last three years.

Faculty were asked to rate the quality of the faculty in the other departments in their field in the study using exactly the same procedure used by Roose and Andersen (1970).

Other variables based on faculty responses included:

Compatibility of Work Environment Scale, a six-item scale with a coefficient alpha of .77 designed to measure the degree of stress and the level of morale of faculty.

Rating of the Quality of Faculty-Student Relations, the mean departmental rating based on faculty ratings on a four-point scale.

Rating of the Humaneness of the Environment, the mean departmental rating based on faculty ratings on a four-point scale.

In addition to the faculty survey, graduate students and alumni from the program were surveyed.

B. Student samples and variables. Returned questionnaires were obtained from 791 chemistry, 893 history, and 967 psychology graduate students. The variables based on their responses that are reported in this paper include:

Quality of Teaching scale, a seven-item scale with a coefficient alpha of .83, designed to measure students' perceptions of a variety of faculty behaviors associated with good teaching.

Quality of the Environment for Learning Scale, a six-item scale with a coefficient alpha of .73, designed to assess the supportiveness and fairness of the department's environment.

Faculty Concern for Students Scale, a seven-item scale with a coefficient alpha of .80 designed to measure students' perceptions of the faculty's accessibility, interest, and helpfulness.

Satisfaction with the Program Scale, a four-item scale with a coefficient alpha of .87 designed to assess overall satisfaction and evaluation of the program.

Quality of the Curriculum Scale, a seven-item scale with a coefficient alpha of .80, designed to assess students' evaluations of the flexibility, depth, and breadth of the curriculum.

Relevance of Degree Requirements Scale, a seven-item scale with a coefficient alpha of .72, designed to assess students' evaluations of the appropriateness and relevance of various degree requirements.

Strength With Which Student Would Advise a Friend to Come to the Department, a rating made on a four-point scale.

C. Alumni data. Questionnaires from 430 chemistry, 349 history, and 393 psychology doctoral recipients were returned. The variables based on their responses include:

Satisfaction With Dissertation Experience Scale, an eleven-item scale with a coefficient alpha of .85 designed to measure alumni satisfaction with a variety of aspects of the dissertation experience.

Value of the Program for Present Work Rating, the mean rating across seven aspects of the doctoral program.

Average Annual Number of Articles and Book Reviews, a self-report measure of alumni scholarly productivity.

Strength With Which Alumni Would Advise Friends to Go to the Department, a rating made on a four-point scale.

Analyses. Departments were grouped into three "quality" groups based on their reputational ratings: high, medium, and low. The publications of faculty in each group as measured in the three areas listed above, for the entire career and for the last three years, was calculated and one-way analysis of variance across the groups performed. Distribution analyses were also made. The mean reputational rating of each department and the mean number of publications on each index was also calculated, and these means correlated to further examine the relationship. Each of these means was also correlated with the means on the other variables just described.

Results

As shown in Table 1, the analysis of variance results showed that all four publications measures involving articles, book chapters, and reviews differed significantly across the three reputational ratings groups for all three fields. (The measures of article and book chapter production are designed to assess research productivity: the measures including book reviews are designed to assess overall scholarly productivity). The trends are most clear for the measures covering the entire career, where the high rated departments produced approximately twice as much per faculty member as the low rated departments, but are much less clear for the measures over the last three years. In fact, for history, the least productive group on measure three were faculty departments rated in the middle rather than those rated low. In

Table 1

Productivity of Faculty in Departments Rated Low, Medium, and High in "Quality": Anova Results

		Means					
		Total Group	Rating Group Low	Medium	High	F	P
1. Articles and Book Chapters: Entire Career	C	54.2	29.2	51.6	71.5	17.9	<.001
	H	11.9	9.5	10.7	14.7	6.6	.002
	P	27.1	17.1	25.2	32.6	12.6	<.001
2. Articles, Book Chapters & Reviews: Entire Career	C	56.8	30.8	54.4	74.3	17.5	<.001
	H	29.4	21.8	27.6	35.3	6.2	.002
	P	29.8	18.5	27.8	35.6	12.4	<.001
3. Books and Monographs: Entire Career	C	1.5	1.3	1.5	1.7	.3	NS
	H	3.4	3.2	3.1	3.9	2.1	NS
	P	2.3	1.3	1.8	3.3	10.4	<.001
4. Articles and Book Chapters: Last Three Years	C	11.2	5.7	10.5	15.2	24.9	<.001
	H	3.0	3.0	2.6	3.4	4.3	.014
	P	6.7	5.1	6.6	7.3	4.8	.009
5. Articles, Book Chapters & Reviews: Last Three Years	C	11.8	6.1	11.3	15.6	24.0	<.001
	H	7.0	5.9	6.6	8.0	4.3	.014
	P	7.3	5.5	7.4	7.8	4.6	.011
6. Books and Monographs: Last Three Years	C	.5	.5	.5	.6	.1	NS
	H	1.0	1.1	.9	1.1	1.9	NS
	P	.8	.5	.8	1.0	3.4	.032

C = Chemistry
H = History
P = Psychology

addition, t-tests showed that the differences between psychology departments rated in the middle and departments rated high were not significant for the last three measures.

The measures of book and monograph production were not significantly different across faculty in departments with different ratings for chemistry and history. In addition, as just noted, the differences between publication rates of psychology faculty in departments rated high and medium on measure six were not significantly different. Thus, book production does not appear to be particularly related to reputational ratings.

To examine the possibility that the high average publication levels of the high rated departments were due to a few very productive faculty, an analysis of the distribution of faculty publication rates in each group was made. There was no sign of concentrations of extraordinarily high producers: the higher productivity of the high rated departments was uniform across the distributions.

The correlations of departmental mean publication indices with the mean reputational ratings and with each other are shown in Table 2. The level of these correlations generally suggests a moderate relationship between reputation and productivity. The only exception was that article and book review production was highly correlated with ratings in chemistry. The correlations also suggest that publication rates per faculty member of journal articles and books are typically only moderately related, if at all.

Table 2

Intercorrelations of Mean Publication Rates and Reputational Ratings

	(1)	(2)	(3)	(4)	(5)
Chemistry					
(1) Articles, Chapters and Book Reviews: Career	1.00				
(2) Articles, Chapters and Book Reviews: Last Three Years	.59*	1.00			
(3) Books, Monographs: Career	.34	.10	1.00		
(4) Books, Monographs: Last Three Years	.23	.18	.74*	1.00	
(5) Reputational Ratings	.85**	.90**	.34	.19	1.00
History					
(1) Articles, Chapters and Book Reviews: Career	1.00				
(2) Articles, Chapters and Book Reviews: Last Three Years	.47*	1.00			
(3) Books, Monographs: Career	.66**	.21	1.00		
(4) Books, Monographs: Last Three Years	.24	.23	.52**	1.00	
(5) Reputational Ratings	.55**	.50*	.52**	.37	1.00
Psychology					
(1) Articles, Chapters and Book Reviews: Career	1.00				
(2) Articles, Chapters and Book Reviews: Last Three Years	.50**	1.00			
(3) Books, Monographs: Career	.43*	.15	1.00		
(4) Books, Monographs: Last Three Years	.34	.26	.70**	1.00	
(5) Reputational Ratings	.33	.47*	.57**	.40*	1.00

*p < .05

**p < .01

It is important to remember that correlations can be misleading, especially when the goal is to evaluate individual departments. For this reason scatterplots between mean departmental publication rates per faculty member and reputational ratings were produced. An example for psychology is given in Figure 1. Here, with departments shown by a code letter, it is clear that the overall correlation is misleading. Department I is the most productive, but is rated lower than Department K, the third least productive. Departments M and H were approximately equally productive, but M obtained the second highest rating, and H the third lowest. Another example, in this case for history, is shown in Figure 2. The figure plots peer ratings of history faculty horizontally and average annual publications vertically, and again, individual programs are represented by randomly assigned letters. Lines have been drawn to indicate the mean of each variable. Although the correlation in Figure 2 is fairly high (.55), the locations of Departments J, C, K, X, and P provide a good illustration of the way in which correlations can be misleading for individual program review. The faculty of Department J, the most productive of all, has one of the lowest ratings, and the faculty of Department C, which has a rating just below that of Department K, averaged nearly twice as many articles per year as K. Department X and P are very similar in productivity, but P is the highest rated program and X is rated below the mean.

In order to better understand the meaning of ratings and publication indices the mean scores for the variables described in the method sections were calculated for each department and correlated

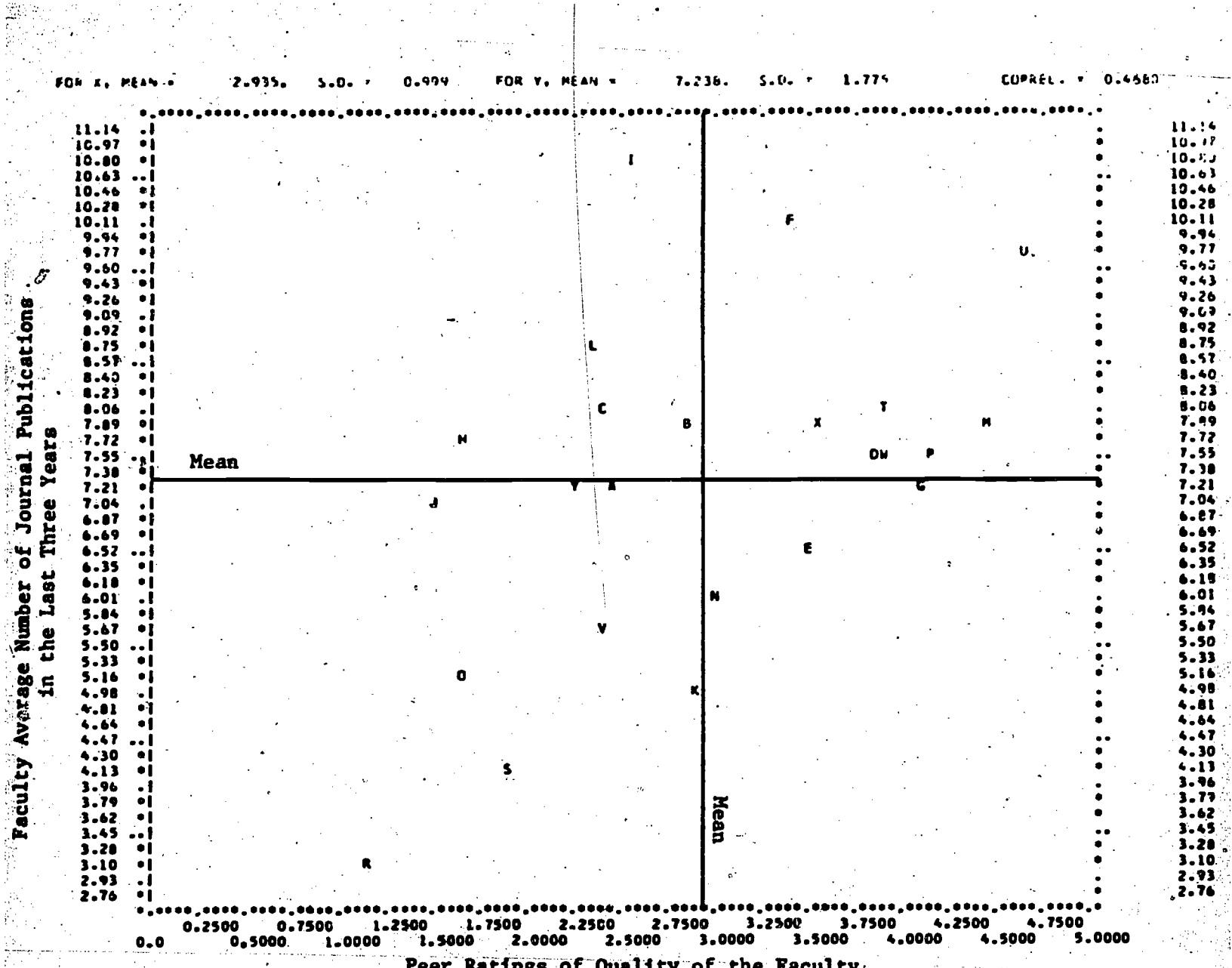

Figure 1. Scatterplot of Peer Ratings of Faculty Quality and Journal Publications in the Last Three Years in Psychology

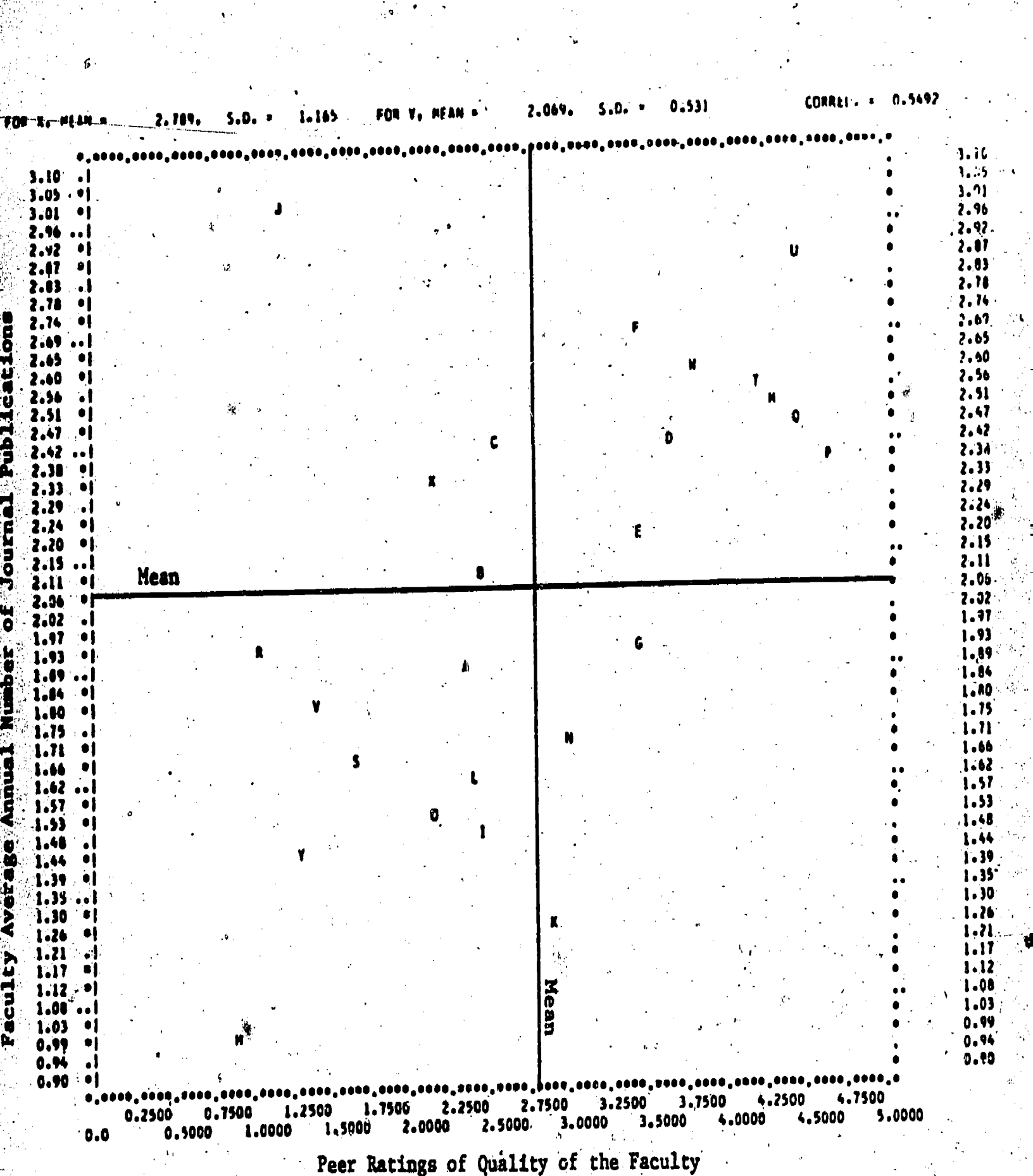

Figure 2. Scatterplot of Peer Ratings of Faculty Quality and Faculty Average Annual Journal Publications in History

with the mean reputational rating of the department and the mean number of journal articles published by the faculty of each department in the last three years. The results are shown in Table 3.

In chemistry ratings were unrelated to all the other indicators of educational quality save one. In history they were negatively related to two and positively related to two. Finally, in psychology ratings were positively related to eight indicators. In contrast, publications were positively related to three indicators in chemistry, none in history, and one in psychology. There were no consistent patterns across fields; none of the other indicators were significantly related to ratings or publications across all three fields, and only two--the faculty Compatibility of Work Environment Scale and the alumni's recommendation to friends to come to the department--were related in two fields, both to ratings.

Although these results are quite complex, there appear to be several trends. First, student indicators of the quality of human relationships are negatively related to ratings in history. (And the positive relationship of ratings to alumni's recommendation of the department to friends is counterbalanced by a near zero relation to a similar recommendation by currently enrolled students). Second, several other student indicators of the quality of the overall academic program are positively related to ratings in psychology. Third, average publication rates did not seem to be especially related to most of the measures. Although it is tempting to speculate on the reasons for these trends, and the differences among fields, it seems clear that, overall, ratings are more strongly

Table 3

Correlations of Reputational Ratings and Publications with Other Indicators of Educational Quality

	Reputational Ratings			Journal Publications Last Three Years		
	C	H	P	C	H	P
From Graduate Student Data						
1. Quality of teaching scale	.09	.00	.52*	-.03	.14	.27
2. Quality of environment for learning scale	-.13	-.48*	.02	-.16	-.25	.15
3. Degree of faculty concern for students scale	-.15	-.39*	.08	-.13	-.21	-.14
4. Satisfaction with program scale	.36	.24	.64*	.27	.16	.27
5. Quality of the curriculum scale	.18	.01	.67**	.09	.07	.33
6. Relevance of degree requirements scale	.02	.19	.68**	-.18	.10	.59*
7. Strength with which students would advice friends to come to department	.19	.02	.46*	.11	.07	.22
From Faculty Data						
8. Compatability of work environment scale	.42*	.56**	.31	.40*	.19	.08
9. Rating of the quality of faculty-graduate student relations	.21	-.09	-.02	.13	-.11	.25
10. Rating of the humaneness of the department's environment	.19	.12	.16	.12	-.15	.25
From Alumni Data						
11. Satisfaction with dissertation experience scale	.14	.28	.57*	.41*	.28	.29
12. Value of academic program for present work rating	-.02	.38	.67**	.00	.24	.23
13. Average annual number of articles and book reviews	.37	-.06	.37	.46*	-.17	.34
14. Strength with which alumni would advise friends to come to the department	.35	.65**	.47*	.28	.42	.36

*p < .05
**p < .01

related to student and alumni descriptions and evaluations of the programs than are faculty publication rates in psychology, but that neither measure is consistently related to other indicators of educational quality in chemistry and history.

Discussion

These analyses showed statistically significant relationships between departmental reputational ratings and a variety of publication measures. However, the results also showed that the ratings tend to reflect the career publications of the faculty more than their recent publications. The relationship as shown in the correlations was also far from perfect, except in chemistry. And, finally, as shown in the scatterplots, the ratings do not reflect the productivity of individual departments at all well. Thus, the utility of the ratings as an indicator of the scholarly contribution of single departments is questionable.

These results also indicate that any particular publication measure should not be used as a measure of overall departmental "quality." The correlation results showed that the different measures of productivity were imperfectly related; some departments that were high on book production were low on journal production. Using only book or journal production would present a false picture of the contributions of individual departments. (The use of publication measures becomes even more problematical when the "quality" of the publications is taken into account). However, more important are the analyses that show both reputational ratings and publication

figures typically to be unrelated to the quality of teaching as reported by students, alumni ratings of the quality of their preparation, faculty concern for students, etc., at least in departments of chemistry and history.

Reputational ratings were related to some of these measures in psychology, but a measure of faculty publication was not. This result suggests that faculty productivity of articles and books may be an adequate measure of one aspect of a department's contribution to the field of psychology, but it is only one aspect of many. Further, although ratings were significantly related to some indicators of the educational excellence of the department, they were not related to others, particularly those reflecting the quality of student-faculty relationships and the overall environment for learning, which suggests that these variables can vary with considerable independence. It has been argued elsewhere (e.g., Hartnett, Clark, & Baird, 1978) that what are needed are multiple indicators of the multiple aspects of departmental educational and scholarly excellence. This would allow one to construct a "profile" of departmental strengths and weaknesses (Clark, 1980) which seems fairer and more sound than attempting to evaluate a department's contribution with one or two indicators. This is especially true in the current period when the amount of financial support for departments is under careful scrutiny and, in some cases, when the continued existence of departments is being questioned.

References

Clark, M. J. The meaning of quality in graduate and professional education. Pp. 85-104 in Katz, J., & Hartnett, R. T. (Eds.), Scholars in the making, Cambridge, Mass.: Ballinger, 1976.

Clark, M. J. Developing and using the graduate program self-assessment service. Educational Testing Service (forthcoming).

Clark, M. J., & Centra, J. A. Determinants of the early career attainments of Ph.D.s. GRE Report (forthcoming).

Clark, M. J., Hartnett, R. T., & Baird, L. L. Assessing dimensions of quality in doctoral education: A technical report of a national study in three fields. Princeton, N. J.: Educational Testing Service, 1976.

Cox, W. M., & Catt, V. Productivity ratings of graduate programs in psychology based on publication in the journals of the American Psychological Association. American Psychologist, 1977, 32, 793-813.

D'Amato, M. R. Kris Foxcat's folly: A fable. American Psychologis 1978, 33, 698-699.

Hartnett, R. T., Clark, M. J., & Baird, L. L. Reputational ratings of doctoral programs. Science, 1978, 199, 1310-1314.

Jones, L. V. The assessment of scholarship, pp. 1-20 in Loveland, E. H. Measuring the hard to measure: New directions for program evaluation, No. 6. San Francisco: Jossey-Bass, 1980.

Kleinmuntz, B. Errors in 1973 biographical directory. American Psychologist, 1978, 33, 693-694. (Comment)

Levin, J. R., et al. University productivity rankings: A psychologist by any other name. American Psychologist, 1978, 33, 693-695.

Robey, J. S. Departments: reputation versus productivity. P.S. 1979, 12, 202-209.

Roose, K. D., & Andersen, C. J. A rating of graduate programs. Washington, D. C.: American Council on Education, 1970.

Ross, A. O. More mischief. American Psychologist, 1978, 33, 696-697.